

# Improved mixing rates of directed cycles by added connection

Balázs Gerencsér\*

Julien M. Hendrickx†

January 10, 2017

## Abstract

We investigate the mixing rate of a Markov chain where a combination of *long distance edges* and *non-reversibility* is introduced: as a first step, we focus here on the following graphs: starting from cycle graph, we select random nodes and add all edges connecting them. We prove a square factor improvement of the mixing rate compared to the reversible version of the Markov chain.

*Keywords:* mixing rate, random graph, non-reversibility.

## 1 Introduction

We study the mixing properties of certain Markov chains which describe how fast the distribution of the state approaches the stationary distribution regardless of the initial conditions. The overall goal is provide significant improvement of mixing by minor modifications of the Markov chain.

Mixing time and rate are fundamental quantities in the study of Markov chains [9], [12] and are the object of active research; they are also highly relevant to applications where mixing properties are strongly tied with performance metrics. This is for example the case for Markov chain Monte Carlo methods, which provide cheap approximation for integrals, and also allow sampling from complex distributions that would otherwise be hard to generate directly [7]. Markov Chains also provide a powerful scheme for approximating the volumes of high dimensional convex bodies [10], [11]. A different application, average consensus, involves the distributed computation of the average of initial values at different agents in a multi-agent system (values which might correspond to measurements, opinions, etc.) [16]. This can be achieved using a Markov chain for which the stationary distribution is uniform. The initial values can be viewed as a probability distribution scaled by a constant, and the Markov chain will approach the uniform distribution multiplied by the same constant, therefore the average of the initial values to be present at each node. Efficient consensus and average consensus approaches actually also play a role in several recent distributed optimization algorithms [13], [14]. For all these applications good performance is crucial and it is determined by the dynamics of the underlying Markov chain. Mixing properties are formulated exactly to answer such questions, which are the topic of the current paper.

There are several ways of obtaining good mixing performances. In many applications, the graph of possible transition is determined by the problem definition, but the specific transition probabilities can be chosen. When these are required to satisfy some strong symmetry properties (reversibility, described later in detail) choosing those optimizing the mixing rate can be formulated as an SDP problem [2], [3], which can be solved numerically using standard methods. Departing from these symmetry properties brings strong technical challenge, at the same time it can actually lead to significant improvement; the mixing time can indeed in some case drop to its square root [4], [12]. On the other hand, when it is possible to modify the

\*B. Gerencsér is with MTA Alfréd Rényi Institute of Mathematics, Hungary, [gerencser.balazs@renyi.mta.hu](mailto:gerencser.balazs@renyi.mta.hu) He is supported by NKFIH (National Research, Development and Innovation Office) grant PD 121107. This work has been carried out during his stay at Université catholique de Louvain, Belgium.

†J. M. Hendrickx is with ICTEAM Institute, Université catholique de Louvain, Belgium, [julien.hendrickx@uclouvain.be](mailto:julien.hendrickx@uclouvain.be) The work is supported by the DYSCO Network (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian Federal Science Policy Office, and by the Concerted Research Action (ARC) of the French Community of Belgium.

graph of possible transitions, astonishing speedup can also be obtained by adding even a small number of randomly selected edges [1], [5], [8].

Our goal in this paper is to study the speedup that can be achieved by adding a small number of random edges *and* making the chain extremely non-reversible. We start with cycle graph of  $n$  nodes, select a lower number  $k$  of them to become hubs, then add extra edges between the hubs. This scheme is motivated by one of the renown models to represent Small World Networks, the Newman-Watts model [5], [15]. The cycle presents a natural way of including asymmetry by introducing a *drift* meaning increased clockwise transition probabilities along the cycle and decreased counter-clockwise ones. At this stage the model needs three parameters to be specified: the placement of the hubs, the added interconnection structure on them, and the asymmetry introduced along the cycle.

In this paper we consider the model where hubs are chosen randomly, all edges between hubs are included and asymmetry is taken to the extreme: the Markov chain is a pure drift along the cycle taking deterministic clockwise steps (except at the hubs). To better understand the dynamics of the process, observe that the state of the Markov chain can be described by an arc (as the cycle is split by the hubs) and the position within that arc. The main challenge here is to show mixing happens both in term of arcs and in terms of positions within. We claim that reaching a mixing rate of  $\Omega(k/n)$  up to  $\log n$  factors is possible and we prove it for  $k = n^\sigma$  where  $0 < \sigma < 1$ .

In comparison, if we were to put pure drift along the cycle but with equidistant hubs, we would have rapid mixing in term of the arc (a perfect one after leaving the first arc), but no mixing at all in term of the position on the arc. Even by decreasing the drift or changing the interconnection structure, the mixing rate will remain  $O((k/n)^2)$  [6].

Furthermore, if we were to stay with the classical, symmetrical transitions along the cycle, the mixing rate will be again  $O((k/n)^2)$ : for an arc at least  $n/k$  long even the hitting time of the ends from the middle is  $\Omega((n/k)^2)$ . This holds for any hub placement and interconnection structure.

After all, we want to emphasize that a speedup with a mixing rate of  $\Omega(k/n)$  is feasible only now that both random hubs and a drift along the cycle are implemented.

The rest of the paper is organized as follows. In Section 2 we formally describe the random graph model and Markov chain on which we focus. In Section 3 we prove the main mixing rate result for a proxy graph model. We then translate our result to the primary graph model in Section 4. In Section 5 simulations are presented complementing our asymptotic analytical results. We also demonstrate how the mixing rate changes when the drift is decreased for the model, suggesting that further performance improvements might be possible. We draw conclusions and outline possible future research directions in Section 6.

## 2 Graph models, Markov chains and mixing rates

The concept of the graphs we consider is the following. We start with a cycle with  $n$  nodes, and randomly select a low number of vertices,  $n^\sigma$  out of the total of  $n$  for some  $0 < \sigma < 1$ , which become hubs. Then we connect all hub nodes with each other. Let us now present the precise definitions.

**Definition 1.** Given  $n, k \in \mathbb{Z}^+$  we define the random graph distribution  $B_n(k)$  as follows. Starting from a cycle graph on  $n$  nodes, we randomly uniformly and independently select  $k$  different edges which we remove. For the  $i$ th remaining arcs,  $1 \leq i \leq k$ , we mark the clockwise endpoint as  $a_i$  and the other end as  $b_i$ . Then we add all edges  $(b_i, a_j)$ , for all  $1 \leq i, j \leq k$ . An example is given in Figure 1.

We introduce an alternative model which reflects approximately the same concept but is technically more convenient due to the added independence.

**Definition 2.** Given  $L \in \mathbb{R}^+$ ,  $k \in \mathbb{Z}^+$ , we define the random graph distribution  $B(L, k)$ . Let us take independently the  $k$  random variables

$$L_i \sim \text{Geo}(1/L), \quad i = 1, \dots, k,$$

where  $\text{Geo}(p)$  denotes a geometric random distribution with parameter  $p$  (and 1 as the smallest possible value). We begin with a graph that is the disjoint union of  $k$  arcs, paths of length  $L_i$ , each of which has a “start point”  $a_i$  and an “end point”  $b_i$ . Then we add all edges  $(b_i, a_j)$ , for  $1 \leq i, j \leq k$ .

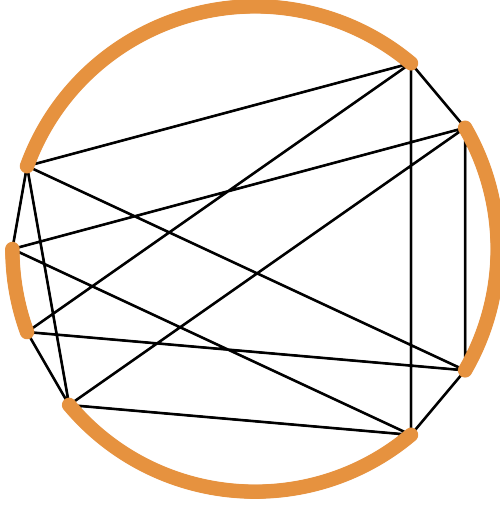


Figure 1: Example graph from  $B_n(k)$  (see Definition 1)

Based on the uniform nature of both constructions we see the following relation between the two models:

**Proposition 3.** *Given  $n, k \in \mathbb{Z}^+$  consider the random graphs  $B(n/k, k)$ . Conditioning on the event of having exactly  $n$  nodes, we get the distribution of  $B_n(k)$ .*

We are interested in the mixing behavior of Markov chains on these graphs. A Markov chain is *reversible* if for any edge  $(u, v)$  of the graph the probability of the  $u \rightarrow v$  transition is the same as the  $v \rightarrow u$  transition. In this paper we go further from the comfortable domain of reversible Markov chains, let us now introduce the ones we will focus on.

**Definition 4.** For any graph coming from  $B_n(k)$  or  $B(L, k)$  we define the *pure drift Markov chain* as follows. Within any arc we set transition probabilities to 1 along the arc all the way from  $a_i$  to  $b_i$ . From any  $b_i$ , we set transition probabilities to  $1/k$  towards all  $a_j$ . A part of such a chain is visualized in Figure 2.

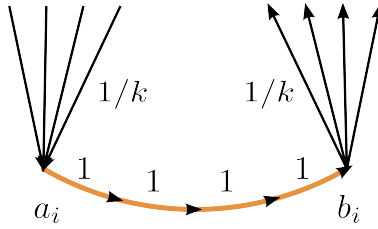


Figure 2: Example arc from the pure drift Markov chain (see Definition 4)

This is a Markov chain which has a doubly stochastic transition matrix, therefore the stationary distribution is uniform. We want to quantify the asymptotic rate as a starting distribution approaches the stationary distribution. This mixing performance of the Markov chain will be measured by the mixing rate:

**Definition 5.** For a Markov chain with transition matrix  $P$  we define the *mixing rate*  $\lambda$  as

$$\lambda = \min \{1 - |\mu| : \mu \neq 1 \text{ is an eigenvalue of } P\}.$$

The natural question arises, why do we fix the transition probabilities to be 1 along the arcs? This excludes mixing to happen happening when the Markov chain is moving along the arcs. There are multiple reasons to do that: First, this demonstrates the power of just the added edges combined with the various

arc lengths, as it already provides fast mixing not only among, but also within the arcs! Second, it is more convenient to analyze such a pure drift version.

Nevertheless, it is normal to ask what happens when there is some diffusion introduced along the arcs. The magnitude of the mixing rate will not change substantially as it is already almost optimal for the pure drift Markov chain (up to logarithmic factors). Still, some improvement does happen, as it will be shown by simulations in Section 5.

We now focus on estimating the mixing rate of the pure drift Markov chain for  $B(L, k)$ . Afterwards, we will transfer these results for the other model  $B_n(k)$ .

### 3 Random polynomials for pure drift Markov chains

In order to find the mixing rate of a Markov chain, we have to know the eigenvalues of its transition matrix. For the current case we transform this eigenvalue problem into finding the roots of a certain polynomial.

**Proposition 6.** *Let us consider the pure drift Markov chain on a random graph from  $B(L, k)$ . Define also*

$$q(z) = \sum_{i=1}^k z^{-L_i}. \quad (1)$$

*Then,  $\mu \in \mathbb{C}$  is an eigenvalue of the transition matrix  $P$  iff  $q(\mu) = k$ .*

*Proof.* Assuming  $\mu$  is the eigenvalue of the transition matrix  $P$  let us find the corresponding eigenvector  $x$ . Observe that each  $a_i$  has incoming edges from exactly the same nodes and the same weights, so the eigenvector must take the same value  $x_{a_i} = h$  at each of them for some  $h$ .

Along the arcs, for two subsequent nodes  $p, p^+$ , from  $xP = \mu x$  we get

$$x_p = \mu x_{p^+}.$$

This implies that along the arcs we see the values  $h, h\mu^{-1}, \dots, h\mu^{-L_i+1}$ . This already completely determines  $x$  up to scaling and ensures the eigenvalue equation for all nodes except the  $a_i$ . We get a valid eigenvector if the equation also holds for  $a_i$ , which takes the form

$$\sum_{i=1}^k \frac{1}{k} h \mu^{-L_i+1} = \mu h,$$

which is equivalent to  $q(\mu) = k$ .

For the other direction, given a  $\mu$  such that  $q(\mu) = k$ , we can again build  $x$  by setting  $1, \mu^{-1}, \dots, \mu^{-L_i+1}$  on each arc, and this will clearly be an eigenvector of  $P$  with eigenvalue  $\mu$ .  $\square$

By Definition 5, the mixing rate is high if the transition matrix  $P$  has no eigenvalue near the complex unit circle. Therefore to get a lower bound on the mixing rate we have to exclude a ring shaped domain for the eigenvalues. The region to be avoided is

$$R_\gamma = \left\{ z : 1 - \frac{1}{L \log^\gamma k} \leq |z| \leq 1, z \neq 1 \right\}, \quad (2)$$

where  $\gamma$  is a parameter to be chosen later. We will show that asymptotically almost surely (a.a.s.) no eigenvalue of  $P$  falls in  $R_\gamma$ . The width of the ring should be viewed as follows. We assume  $L$  and  $k$  are of similar magnitudes meaning that they have a polynomial growth rate w.r.t. each other. Therefore the width is at most a logarithmic factor lower than  $1/L$ . Our key result is the following:

**Theorem 7.** *Assume  $k, L \rightarrow \infty$  while  $\rho_l < \log L / \log k < \rho_u$  for some constants  $0 < \rho_l < \rho_u < \infty$ , and fix  $\gamma > 4$ . Based on the graph model  $B(L, k)$  we use  $q(z)$  from (1) and  $R_\gamma$  from (2). Then for any  $c, d > 0$  we have*

$$P(\exists z \in R_\gamma, q(z) = k) = O(k^{-c} L^{-d}).$$

*In particular, for the mixing rate we get  $\lambda > 1/(L \log^\gamma k)$  a.a.s.*

Note that the right hand side can be simplified a bit. The relation between  $L$  and  $k$  ensures  $k^{\rho_l} < L < k^{\rho_u}$  after a while. Therefore the  $L$  term on the right hand side can be replaced by a power of  $k$ , and it is sufficient to show that the probability in question is  $O(k^{-c})$  for all  $c > 0$ .

We show the claim in four steps. First we ensure that we can assume the  $L_i$  variables to be bounded with high probability, this will make further estimates possible. We then check  $z$  coming from different parts of  $R_\gamma$ : we start with positive reals, then we treat complex numbers in two different ways depending on their arguments.

Intuitively the reason is the following. For some real  $0 < z < 1$ ,  $q(z)$  will be too large. Next, take  $z$  with low arguments, now all  $z^{-L_i}$  will be in the same half-plane giving a non-zero imaginary part to  $q(z)$ . When the argument is far enough such that  $q(z)$  has a chance to have zero imaginary part again, the  $z^{-L_i}$  will point in so many different directions that the cancellations will force the real part below  $k$ , a.a.s.

Now let us make all this precise. We will confirm that each of the intuitive steps above work with high probability, whenever the  $L_i$  are different enough and are not extremely large. We then join these steps to give a proof of the theorem.

The probabilistic upper bound we need on the  $L_i$  can be formulated in the following way:

**Lemma 8.** *For any  $C > 1$ , there holds*

$$P(\max_i L_i \geq CL \log k) = O(k^{1-C}).$$

*Proof.* Remember that  $L_i$  are i.i.d. variables with law  $\text{Geo}(1/L)$ . Therefore we have that

$$P(L_i \geq CL \log k) = P(L_i \geq \lceil CL \log k \rceil) = \left(1 - \frac{1}{L}\right)^{\lceil CL \log k \rceil} \leq \left(1 - \frac{1}{L}\right)^{CL \log k}.$$

Knowing  $(1 - 1/L)^L < 1/e$  we get

$$P(L_i \geq CL \log k) \leq e^{-C \log k} = k^{-C}.$$

To treat all  $L_i$  together for  $1 \leq i \leq k$ , we use a simple union bound.

$$P(\max_i L_i \geq CL \log k) \leq k^{1-C}.$$

□

From now on, we will only investigate the (a.a.s.) event that the maximal  $L_i$  is small as shown in Lemma 8. Let us call this event  $S(C)$ . We now check  $z$  coming from different parts of  $R_\gamma$ . The simplest case is when  $z$  is a positive real:

**Lemma 9.** *Assume  $z \in (0, 1)$ . Then  $q(z) \neq k$ .*

*Proof.* In this case,  $q(z)$  is composed of  $k$  positive terms, each of them being larger than 1. Consequently the sum is higher than  $k$ . □

Next we show that there is no  $z \in R_\gamma$  with small arguments for which  $q(z) = k$ .

**Lemma 10.** *Assume  $S(C)$  and take  $z \in R_\gamma$  such that  $0 \neq |\arg(z)| < \pi/(CL \log k)$ . Then  $\Im q(z) \neq 0$ . Consequently  $q(z) \neq k$ .*

*Proof.* Without loss of generality, assume  $\arg(z) > 0$ . The event  $S(C)$  ensures  $L_i < CL \log k$ , so  $z^{-L_i}$  will all be in the same half-plane, they will have an argument in  $(-\pi, 0)$ . For all these values, the imaginary part is negative, and so is their sum. Therefore it simply cannot be 0. □

It remains to check the elements of  $R_\gamma$  whose argument is “large”. The arguments in question are those in

$$A := \left[ \frac{\pi}{CL \log k}, 2\pi - \frac{\pi}{CL \log k} \right]. \quad (3)$$

We argue that the angles of  $z^{-L_i}$  become so different that strong cancellations will happen. Our main proposition formally stating this is the following:

**Proposition 11.** Choose constants  $\alpha, \beta > 1$  and also  $\rho_l, \rho_u$  as in Theorem 7, and require  $\rho_l < \log L / \log k < \rho_u$ . We use  $k, L, L_i$  as in Definition 2. We define

$$m = \log^\alpha k, \quad \delta = \log^{-\beta} k.$$

Then for  $k, L$  large enough we have

$$P \left( \sup_{x \in A} \sum_{i=1}^m \cos^+(L_i x) < m - \delta^2 \right) \geq \frac{1}{3},$$

where  $\cos^+(y) = \max(\cos(y), 0)$ .

*Proof.* Broadly speaking we want to have one of the  $L_i x$  terms to be far from  $2k\pi$ , which should decrease enough the sum of the cosines. For a single  $x$ , we state the following lemma.

**Lemma 12.** Use  $k, L, L_i$  as in Definition 2. Fix  $\beta > 1$ ,  $x \in A$  and choose  $D \subset [0, 2\pi]$  which is a modulo  $2\pi$  interval with  $|D| = 6\delta = 6\log^{-\beta} k$ . Then for  $k, L$  large enough we have

$$P(\{L_1 x\} \in D) \leq \frac{3}{4},$$

where  $\{a\}$  stands for  $a \bmod 2\pi$ .

*Proof.* Each element of the series  $\{x\}, \{2x\}, \{3x\}, \dots$  is either in  $D$  or not. We can therefore split the series into blocks that are in  $D$  and to blocks that are not. Let  $t_1$  be the first coefficient such that  $\{t_1 x\} \in D$ , then we can define the blocks

$$\begin{aligned} \{t_i x\}, \{(t_i + 1)x\}, \dots, \{(s_i - 1)x\} &\in D, \\ \{s_i x\}, \{(s_i + 1)x\}, \dots, \{(t_{i+1} - 1)x\} &\notin D. \end{aligned}$$

We will show the statement

$$P(L_1 \in [t_i, s_i - 1]) \leq 3P(L_1 \in [s_i, t_{i+1} - 1]), \quad (4)$$

This is sufficient to complete the proof because by summing them all (and also including the numbers  $1, 2, \dots, t_1 - 1$ ) we get

$$P(\{L_1 x\} \in D) \leq 3P(\{L_1 x\} \notin D),$$

and this immediately confirms the lemma.

We now compare the number of elements in the blocks above. We want to relate  $t_{i+1} - s_i$  with  $s_i - t_i$ . For  $k, L$  large enough we have  $|D| = 6\delta < \pi$ . Then we have

$$s_i - t_i \leq \frac{|D|}{\min(x, 2\pi - x)} + 1 \leq \frac{2\pi - |D|}{\min(x, 2\pi - x)} + 1 \leq t_{i+1} - s_i + 1.$$

The  $+1$  term is needed to compensate possible boundary effects. The same claim holds for  $x > \pi$ , it corresponds to moving around the angles in the other direction. Using  $t_{i+1} - s_i \geq 1$  we also have

$$\frac{1}{2}(s_i - t_i) \leq t_{i+1} - s_i. \quad (5)$$

We now show an upper bound on  $s_i - t_i$  by comparing the interval length with the smallest possible step size. Remember that the step size  $x$  must fall in  $A$  defined in (3).

$$s_i - t_i \leq \frac{6\delta}{\min\{\min(x, 2\pi - x) : x \in A\}} + 1 = \frac{6}{\log^\beta k} \cdot \frac{CL \log k}{\pi} + 1 \leq 2CL \log^{1-\beta} k + 1. \quad (6)$$

After deducing inequalities (5), (6) on the number of elements in the blocks, we also compare the probabilities of  $L_1$  falling within them. The probability of  $L_1 = j$  decreases with  $j$ . Therefore the probability of  $L_1$  falling into an interval is smaller than twice the probability of falling into the first half if the interval.

$$P(L_1 \in [t_i, s_i - 1]) \leq P(L_1 \in [t_i, t_i + (t_i - s_i) - 1]) \leq 2P\left(L_1 \in \left[t_i, t_i + \frac{s_i - t_i}{2} - 1\right]\right) = \dots \quad (7)$$

As  $L_1$  is a geometric random variable, shifting the interval of interest by  $s_i - t_i$  introduces only a simple multiplicative factor:

$$\dots = 2\left(1 - \frac{1}{L}\right)^{t_i - s_i} P\left(L_1 \in \left[s_i, s_i + \frac{s_i - t_i}{2} - 1\right]\right) \leq \dots$$

We enlarge the target interval from length  $(s_i - t_i)/2$  to  $t_{i+1} - s_i$  relying on (5). Clearly by this the probability cannot decrease.

$$\dots \leq 2\left(1 - \frac{1}{L}\right)^{t_i - s_i} P(L_1 \in [s_i, t_{i+1} - 1]).$$

For the coefficient at the end of (7) we use (6) to get

$$2\left(1 - \frac{1}{L}\right)^{t_i - s_i} \leq 2\left(1 - \frac{1}{L}\right)^{-2CL \log^{1-\beta} k - 1} \leq 2\left(1 + \frac{1}{L}\right) \exp(3C \log^{1-\beta} k) \leq 3,$$

for  $k, L$  large enough. Here we use that  $(1 - 1/L)^{-2L}$  is approximately  $\exp(2)$ , so it is certainly below  $\exp(3)$  for  $L$  large enough.

Substituting this last bound together into (7) leads to (4). This is enough to complete the proof as we have seen before.  $\square$

To come back to the proof of Proposition 11, let us choose  $D = [-3\delta, 3\delta]$ . Whenever we have  $\{L_i x\} \notin D$  then it implies  $\cos^+(L_i x) < 1 - 2\delta^2$ . Using Lemma 12 and knowing that the  $L_i$  are independent random variables we have

$$P\left(\sum_{i=1}^m \cos^+(L_i x) \geq m - 2\delta^2\right) \leq P(\{L_i x\} \in D)^m \leq \left(\frac{3}{4}\right)^m. \quad (8)$$

This is the type of probability bound we are looking for, but only for a single  $x$ . Next we extend it to all  $x \in A$  simultaneously, where  $A$  is the interval of interest of arguments (3). As an intermediate step, take a grid  $x_j$  of resolution  $\epsilon = 2\delta^2/(mCL \log k)$  on  $A$ . Using the union bound for the grid we see

$$\begin{aligned} P\left(\sup_j \sum_{i=1}^m \cos^+(L_i x_j) \geq m - 2\delta^2\right) &\leq \sum_{j=1}^m P\left(\sum_{i=1}^m \cos^+(L_i x_j) \geq m - 2\delta^2\right) \\ &\leq \frac{\pi mCL \log k}{2\delta^2} \left(\frac{3}{4}\right)^m = \frac{\pi}{2} \log^\alpha k CL \log k \log^{2\beta} k \left(\frac{3}{4}\right)^{\log^\alpha k} \\ &= \frac{C\pi}{2} L \log^{2\beta+\alpha+1} k \left(\frac{3}{4}\right)^{\log^\alpha k}. \end{aligned} \quad (9)$$

In this final term  $(3/4)^{\log^\alpha k}$  is super-polynomial in  $k$  as  $\alpha > 1$ . All others have polynomial or lower rate in  $k$ . Consequently we see that the right hand side of (9) will become arbitrarily small as  $k, L$  grows. In particular, it will go below  $2/3$ .

At this point we have the probability estimate for the grid, we need to extend this to the complete interval  $A$ , introduced in (3). We show that for  $k, L$  large enough we have

$$P\left(\sup_{x \in A} \sum_{i=1}^m \cos^+(L_i x) \geq m - \delta^2\right) < P\left(\sup_j \sum_{i=1}^m \cos^+(L_i x_j) \geq m - 2\delta^2\right). \quad (10)$$

Indeed, for any  $x \in A$  there is a grid point  $x_j$  at most  $\epsilon/2$  away. As the derivative of  $\cos^+$  stays within  $[-1, 1]$ , the change of the sum when moving to  $x_j$  from  $x$  is at most

$$\sum_{i=1}^m \frac{\epsilon}{2} L_i \leq m \frac{\delta^2}{mCL \log k} CL \log k = \delta^2.$$

Therefore when the sum on the left hand side of (10) is at least  $m - \delta^2$  for a certain  $x \in A$  then there also must be a grid point  $x_j$  for which the sum is at least  $m - 2\delta^2$ . The inclusion of the events shows the inequality for the probabilities.  $\square$

*Proof of Theorem 7.* Choose  $C \geq c+1$ . Let us assume  $\max_i L_i < CL \log k$ . This not holding is an exceptional event of probability  $O(k^{-c})$  as shown in Lemma 8. In order to exclude the roots from all  $R_\gamma$ , we split this region into three parts.

When  $0 < z < 1$  is a positive real, it cannot be a solution according to Lemma 9. When  $z$  has a small argument, that is,  $|\arg(z)| < \pi/(CL \log k)$ , we refer to Lemma 10 to confirm  $z$  cannot be a root,  $q(z) \neq k$ .

The remaining case is when  $z$  has a large argument, that is,  $\arg(z) \in A$ . We aim to bound  $\Re q(z)$ . On one hand, we estimate the magnitude of the terms  $z^{-L_i}$ . Then we combine these with the cosines of the arguments to find the contribution to the real part of  $q(z)$ . Here we rely on Proposition 11, but let us make this precise.

We need to check the magnitude of the terms  $z^{-L_i}$ . Knowing  $|z| > 1 - 1/(L \log^\gamma k)$  and  $L_i < CL \log k$  for  $k, L$  large enough we have

$$\begin{aligned} |z|^{-L_i} &\leq \left(1 - \frac{1}{L \log^\gamma k}\right)^{-CL \log k} \leq \left(1 + \frac{2}{L \log^\gamma k}\right)^{CL \log k} \\ &\leq \exp(2C \log^{1-\gamma} k) \leq 1 + \frac{4C}{\log^{\gamma-1} k}. \end{aligned}$$

Considering  $\Re q(z)$ , this gives

$$\Re q(z) = \sum_{i=1}^k |z|^{-L_i} \cos(L_i x) \leq \sum_{i=1}^k |z|^{-L_i} \cos^+(L_i x) \leq \left(1 + \frac{4C}{\log^{\gamma-1} k}\right) \sum_{i=1}^k \cos^+(L_i x). \quad (11)$$

Let us arrange the  $k$  elements of this sum into groups of  $m = \log^\alpha k$  elements, consequently resulting in  $k/m$  such groups. Here we omit the minor details dealing with  $m, k/m$  not being integers in general. Let the sum of these groups be  $S_1, S_2, \dots, S_{k/m}$ . According to Proposition 11 we have

$$P(S_i > m - \delta^2) \leq \frac{1}{3},$$

and each of these events are independent. Therefore the number of such events happening follows a  $\text{Binom}(k/m, r)$  distribution for some  $r \leq 1/3$ . From standard Chernoff bounds we see that

$$P\left(\text{Binom}\left(\frac{k}{m}, r\right) > \frac{k}{2m}\right) \leq P\left(\text{Binom}\left(\frac{k}{m}, \frac{1}{3}\right) > \frac{k}{2m}\right) \leq \sqrt{\frac{8}{9} \frac{k}{m}} < k^{-c},$$

for  $k$  large enough. Consequently, at most  $k/(2m)$  of  $S_i$  are large (with the exception of an event with probability of  $O(k^{-c})$ ). In this case we have

$$\sum_{i=1}^k \cos^+(L_i x) = \sum_{j=1}^{k/m} S_j \leq \frac{k}{2m} m + \frac{k}{2m} (m - \delta^2) = k - \frac{\delta^2 k}{2m}.$$

Plugging this back into (11) we arrive at

$$\Re q(z) \leq \left(1 + \frac{4C}{\log^{\gamma-1} k}\right) k \left(1 - \frac{\delta^2}{2m}\right) = k \left(1 + \frac{4C}{\log^{\gamma-1} k}\right) \left(1 - \frac{1}{2 \log^{2\beta+\alpha} k}\right).$$

Let us choose  $\gamma > 2\beta + \alpha + 1$ . With such a choice, the term  $k$  above gets multiplied by a coefficient lower than 1 for  $k$  large enough. This shows  $\Re q(z) < k$  which implies  $z$  is not a solution of  $q(z) = k$ . This a.a.s. holds simultaneously for all  $z \in R_\gamma$ ,  $\arg(z) \in A$ .

Regarding the parameters, previously for Proposition 11 we only needed to ensure  $\alpha, \beta > 1$ . Therefore we can apply Proposition 11 and this reasoning for  $\alpha = \beta = 1 + \epsilon$ ,  $\gamma = 4 + 4\epsilon$  for any  $\epsilon > 0$ , eventually allowing any  $\gamma > 4$ . During the proof, we had two small exceptional events, both having probability  $O(k^{-c})$ . This confirms the theorem with the condition on  $\gamma$  and with the probability bound on the exceptional cases.  $\square$



## 4 Mixing rates for $B_n(k)$

Theorem 7 of the previous section tells about the eigenvalues for the pure drift Markov chain on the graph model  $B(L, k)$ . It guarantees the absence of eigenvalues with large absolute value (except at 1) with high probability. Based on this result we get:

**Theorem 13.** *Assume  $k, L \rightarrow \infty$  while  $\rho_l < \log L / \log k < \rho_u$  for some constants  $0 < \rho_l < \rho_u < \infty$ . Then for any  $\gamma > 4$  a.a.s. we have the following bound on the mixing rate for  $B(L, k)$ :*

$$\lambda > \frac{1}{L \log^\gamma k}.$$

*Proof.* This lower bound is a direct consequence of Theorem 7 and the definitions of  $R_\gamma$  and the mixing rate  $\lambda$ .  $\square$

We now translate this result to the other graph model  $B_n(k)$ , where the total number of nodes are fixed beforehand.

**Theorem 14.** *Assume  $n, k \rightarrow \infty$  while  $\rho_l < \log n / \log k < \rho_u$  for some constants  $0 < \rho_l < \rho_u < \infty$ . Then for any  $\gamma > 4$  a.a.s. we have the following bound on the mixing rate for  $B_n(k)$ :*

$$\lambda > \frac{k}{n \log^\gamma k}.$$

*Proof.* We obtain the model  $B_n(k)$  by starting from  $B(n/k, k)$  and then conditioning on the event that the node count of the graph matches exactly  $n$ . Let us call this event  $M(n, k)$ . We develop the probability of this to happen.

The node count of  $B(n/k, k)$  is calculated by adding up  $k$  independent geometric random variables. This is closely related to the negative binomial distribution, therefore we can interpret it as follows. We perform independent trials of a binary test that has a failure probability  $k/n$ . We wait until the  $k$ th failure and check whether the number of total trials is  $n$ . We arrive in this event if there is a failure exactly at the  $n$ th trial, and we had  $k - 1$  failures before. For any given configuration of  $k$  positions for failures and  $(n - k)$  for successful trials, the probability is

$$\left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k}.$$

We have to include any configuration of the first  $k - 1$  failures (the last one is fixed). Summing these up we get the overall probability

$$P(M(n, k)) = \binom{n-1}{k-1} \left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k} = \binom{n-1}{k-1} \frac{k^k (n-k)^{n-k}}{n^n}. \quad (12)$$

We develop a simple asymptotic estimate for this probability. From the Stirling formula we know that

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n} = 1.$$

For conciseness, we will use the  $\approx$  relation if the ratio of the two quantities is 1 in the limit. In this spirit we get

$$\binom{n}{k} \approx \frac{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n}{\sqrt{2\pi k} \left(\frac{k}{e}\right)^k \sqrt{2\pi(n-k)} \left(\frac{n-k}{e}\right)^{n-k}} = \sqrt{\frac{n}{2\pi k(n-k)}} \frac{n^n}{k^k (n-k)^{n-k}}.$$

Let us plug this back to (12), while noting  $\binom{n-1}{k-1} = \frac{k}{n} \binom{n}{k}$ .

$$P(M(n, k)) \approx \frac{k}{n} \sqrt{\frac{n}{2\pi k(n-k)}} = \sqrt{\frac{k}{2\pi n(n-k)}}.$$

As a very crude bound we get for  $n, k$  large enough that

$$P(M(n, k)) > \frac{1}{n}. \quad (13)$$

We can now use Theorem 7 as the conditions currently imposed imply the conditions of Theorem 7. Therefore we get that the probability of having a root in the forbidden ring is bounded as  $O(L^{-2}k^{-2}) = O(n^{-2})$ . Also, we have just seen in (13) that starting from  $B(n/k, k)$ , we get a graph with  $n$  nodes with probability at least  $1/n$  for  $n, k$  large enough.

Consequently, the probability of having a root in the forbidden ring is negligible even compared to the event  $M(n, k)$ . Conditioning on  $M(n, k)$  means generating an instance of  $B_n(k)$ . We have just seen that the probability of getting a graph with a root in  $R_\gamma$  asymptotically vanishes even compared to this event. This completes the proof.  $\square$

## 5 Simulations

Following the asymptotic theoretical results we perform complementing simulations to analyze the tightness of the bounds obtained. We also explore further numerically for the next step of research that is not yet treated analytically.

The mixing results we have are exciting as we see a strong speedup compared to the similar reversible Markov chain with transition matrix  $\tilde{P} = (P + P^\top)/2$ . By this we set all transition probabilities on all edges to be equal in the two directions. For this Markov chain, if the initial distribution is concentrated in the middle of the longest arc, the Central Limit Theorem ensures that even after  $\Omega(L^2 \log^2 k)$  steps the probability of not leaving the arc is bounded away from 0. Consequently we get a lower bound of the same order for the mixing time and which in turn translates to the mixing rate bound

$$\lambda < \frac{C}{L^2 \log^2 k}, \quad (14)$$

which is a square factor worse than our new results for the non-reversible Markov chain.

Simulations are in line with the speedup we see when comparing (14) with Theorem 13. Figure 3 is a log-log histogram showing the decrease of  $\lambda$  as the node count  $n$  increases. The histogram presents the simulation results for the non-reversible and reversible Markov chain and we do observe the strong separation predicted by the theoretical results. The stripe on the top presents  $\lambda$  for the non-reversible Markov chains while the bottom one corresponds to the reversible ones. Figure 3 is based on 200.000 random Markov chains with  $n$  ranging from 54 to 2980 and with  $k = \lfloor \sqrt{n} \rfloor$ . As we are interested in typical behavior of these randomized Markov chains, we discarded the top and bottom 5% of the results for every  $n$  considered.

The two type of Markov chains we compared can be seen as the extremal setups: either the asymmetry is so strong that steps are deterministic along the cycle, or we have perfect symmetry. There is however a full spectrum of intermediate situations, and one may wonder which level of asymmetry is optimal. We have seen that full asymmetry is better in terms of mixing performance than full symmetry. On Figure 4, starting from the reversible Markov chain, we gradually change the transition probabilities along the cycle until we reach the current extreme asymmetric case. Specifically, for  $1/2 \leq q \leq 1$  we set the transition matrix  $P_q = qP + (1 - q)P^\top$  and compute the mixing rate of the resulting Markov chain. Here we have  $P_{1/2} = \tilde{P}$ ,  $P_1 = P$  as expected. We perform simulations for  $B_{500}(10)$  and  $B_{500}(50)$ . In both case, 8000 random graphs were generated and the mixing rates were computed for all graphs and for  $q$  moving along  $[1/2, 1]$ . Again, the top and bottom 5% were discarded. The means of the resulting mixing rates are presented in Figure 4 together with the standard deviations. The figures show that the optimal choice is near the extremal non-reversible case, confirming our concept. Still, interestingly a minor offset towards the reversible version still increases the mixing rate. Intuitively the two effects of the modification match well: the small loss in the speed of moving along the cycle is well compensated by the local mixing introduced.

The analytic treatment of the intermediate Markov chains for  $q \neq 1/2, 1$  brings new challenges as the non-reversible feature is still present while we lose the deterministic nature of the movement along the cycle.

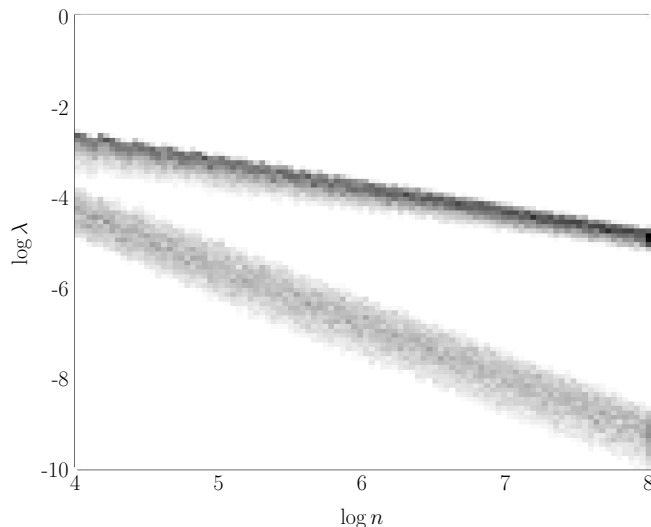


Figure 3: Histograms for the mixing rates  $\lambda$  for the Markov chains on the graphs  $B_n(\lfloor \sqrt{n} \rfloor)$ . The upper stripe corresponds to non-reversible Markov chains while the lower one to the reversible variants.

## 6 Conclusions

We have seen in Theorem 13 and Theorem 14 that for both models  $B_n(k)$  and  $B(L, k)$  the mixing rate of the non-reversible Markov chain considered is much higher than the similar reversible one. The results confirm that the simultaneous application of adding long distance edges and also setting the Markov chain to be non-reversible dramatically improves the mixing rate. We believe this phenomenon is promising and could provide similar speedup effects for other reference graphs, other methods to add random edges and other means of introducing non-reversibility.

We have also seen numerically in Figure 4 that being fully non-reversible is not necessarily optimal in this context, even though it is significantly better than being fully reversible.

Therefore one of the open questions is to find the optimal Markov chain among the intermediate cases. Another goal for future research is to consider the more realistic situation where the hubs do not have such high number of connection, for instance, by replacing the complete graph on the selected  $cn^\sigma$  nodes with a random matching on them. Simulations similar to Figure 3 in [6] suggest that a similar speedup is to be expected.

## References

- [1] L. ADDARIO-BERRY AND T. LEI, *The mixing time of the Newman–Watts small world*, in Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 18 Jan 2012, pp. 1661–1668.
- [2] S. BOYD, P. DIACONIS, P. PARRILO, AND L. XIAO, *Fastest mixing Markov chain on graphs with symmetries*, SIAM J. Optim., 20 (2009), pp. 792–819.
- [3] S. BOYD, P. DIACONIS, AND L. XIAO, *Fastest mixing Markov chain on a graph*, SIAM Rev., 46 (2004), pp. 667–689 (electronic).
- [4] P. DIACONIS, S. HOLMES, AND R. M. NEAL, *Analysis of a nonreversible Markov chain sampler*, Ann. Appl. Probab., 10 (2000), pp. 726–752.
- [5] R. DURRETT, *Random Graph Dynamics*, Cambridge University Press, 2006.

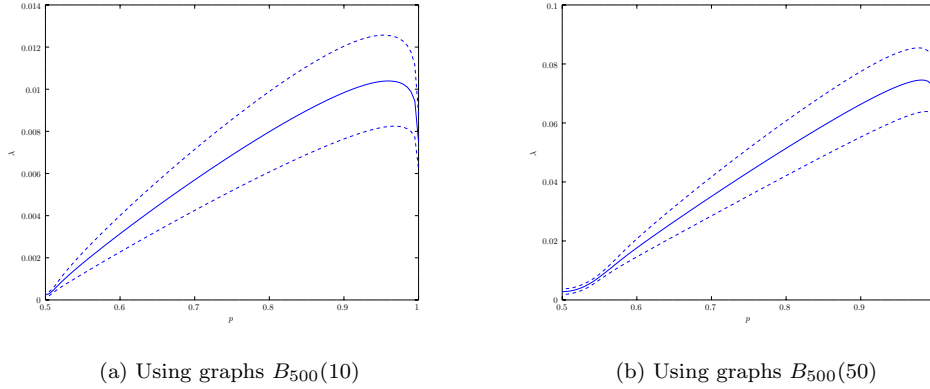


Figure 4: Mixing rates of the transition matrices  $P_q = qP + (1-q)P^\top$  for the interpolation between the pure drift ( $q = 1$ ) Markov chain and its reversible version ( $q = 0.5$ ). Solid lines follow the means while dashed lines indicate the standard deviations around the means.

- [6] B. GERENCSÉR, *Mixing times of Markov chains on a cycle with additional long range connections*, arXiv:1401.1692, 2014.
- [7] M. JERRUM, *Mathematical foundations of the Markov chain Monte Carlo method*, in Probabilistic methods for algorithmic discrete mathematics, vol. 16 of Algorithms and Combinatorics, Springer, 1998, pp. 116–165.
- [8] M. KRIVELEVICH, D. REICHMAN, AND W. SAMOTIJ, *Smoothed analysis on connected graphs*. Preprint, arXiv:1307.4884, 2014.
- [9] D. LEVIN, Y. PERES, AND E. WILMER, *Markov chains and mixing times*, American Mathematical Society, 2009.
- [10] L. LOVÁSZ AND S. VEMPALA, *Hit-and-run from a corner*, SIAM Journal on Computing, 35 (2006), pp. 985–1005.
- [11] ———, *Simulated annealing in convex bodies and an  $o^*(n^4)$  volume algorithm*, Journal of Computer and System Sciences, 72 (2006), pp. 392–417.
- [12] R. MONTENEGRO AND P. TETALI, *Mathematical aspects of mixing times in Markov chains*, Theor. Comput. Sci., 1 (2006), pp. 237–354.
- [13] A. NEDIĆ AND A. OLSHEVSKY, *Distributed optimization over time-varying directed graphs*, IEEE Transactions on Automatic Control, 60 (2015), pp. 601–615.
- [14] A. NEDIC, A. OZDAGLAR, AND P. A. PARRILO, *Constrained consensus and optimization in multi-agent networks*, IEEE Transactions on Automatic Control, 55 (2010), pp. 922–938.
- [15] M. NEWMAN, C. MOORE, AND D. WATTS, *Mean-field solution of the small-world network model*, Phys. Rev. Lett., 84 (2000), pp. 3201–3204.
- [16] A. OLSHEVSKY AND J. N. TSITSIKLIS, *Convergence speed in distributed consensus and averaging*, SIAM J. Control Optim., 48 (2009), pp. 33–55.